regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences. Modification of the sequences encoding the particular protein of interest may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; or to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site which is in reading frame with and under regulatory control of the control sequences.

[0042] Suitable marker sequences for identification and isolation of correctly transfected cells include the thymidine kinase (tk), dihydrofolate reductase (DHFR), and aminoglycoside phosphotransferase (APH) genes. The latter imparts resistance to the aminoglycoside antibiotics, such as kanamycin, neomycin, and geneticin. These, and other marker genes such as those encoding chloramphenicol acetyltransferase (CAT) and β-galactosidase (β-gal), may be incorporated into the primary nucleic acid cassette along with the gene expressing the desired therapeutic protein, or the selection markers may be contained on separate vectors and cotransfected.

[0043] The term "biochemically equivalent variations" means protein or nucleic acid sequences which differ in some respect from the specific sequences disclosed herein, but nonetheless exhibit the same or substantially the same functionality. In the case of cDNA, for example, this means that modified sequences which contain other nucleic acids than those specifically disclosed are encompassed, provided that the alternate cDNA encodes mRNA which in turn encodes a protein of this invention. Such modifications may involve the substitution of only a few nucleic acids, or many. The modifications may involve substitution of degenerate coding sequences or replacement of one coding sequence with another; introduction of non-natural nucleic acids is included. Preferably, the modified nucleic acid sequence hybridizes to and is at least 95% complementary to the sequence of interest.

[0044] Similarly, in the case of the proteins and polypeptides of this invention, alterations in the amino acid sequence which do not affect functionality may be made. Such "biochemically equivalent muteins" may involve replacement of one amino acid with another, use of side chain modified or non-natural amino acids, and truncation. The skilled artisan will recognize which sites are most amenable to alteration without affecting the basic function.

[0045] The expression products described herein are proteins and polypeptides having a defined chemical sequence. However, the precise structure depends on a number of factors, particularly chemical modifications common to proteins. For example, since all proteins contain ionizable amino and carboxyl groups, the protein may be obtained in acidic or basic salt form, or in neutral form. The primary amino acid sequence may be derivatized using sugar molecules (glycosylation) or by other chemical derivatizations involving covalent or ionic attachment with, for example, lipids, phosphate, acetyl groups and the like, often occurring through association with saccharides. These modifications may occur in vitro, or in vivo, the latter being performed by a host cell through post-translational processing systems. Such modifications may increase or decrease the biological activity of the molecule, and such chemically modified molecules are also intended to come within the scope of the invention.

[0046] B. Hypocretin Proteins and Polypeptides

[0047] Hypocretin or clone H35, has been cloned in both rat and mouse. The amino acid residue sequence in these two mammalian species is not identical but is sufficiently similar to permit generalization regarding function, and so that one can identify and isolate the hypocretin gene in any mammalian species.

[0048] Variations at both the amino acid and nucleotide sequence level are described in isolates of hypocretin, and such variations are not to be construed as limiting. For example, allelic variation within a mammalian species can tolerate a several percent difference between isolates of a type of hypocretin, which differences comprise non-deleterious variant amino acid residues. Thus a protein of about 95% homology, and preferably at least 98% homology, to a disclosed hypocretin is considered to be an allelic variant of the disclosed hypocretin, and therefore is considered to be a hypocretin of this invention.

[0049] As disclosed herein, hypocretin is produced first in vivo in precursor form, and is then processed into smaller polypeptides having biological activity as described herein. Insofar as these different polypeptide forms are useful, the term hypocretin protein or polypeptide connotes all species of polypeptide having an amino acid residue sequence derived from the hypocretin gene.

[0050] The complete coding nucleotide sequence, clone 35, of rat H35 cDNA is 569 nucleotides in length, and is listed in SEQ ID NO 3. The complete preprohypocretin cDNA clone presents a 390 nucleotide open reading frame (ORF) plus triplet termination codons (**FIG. 5**). There is a N-terminal signal peptide with a cleavage site between amino acid positions 27 and 28, corresponding to a cleavage site after nucleotide position 172 of SEQ ID NO: 3.

[0051] Translation of this rat cDNA sequence produces a novel protein of 130 amino acid residues, referred to as rat preprohypocretin. The amino acid sequence of rat preprohypocretin is listed in SEQ ID NO: 1. The amino acid sequence of mouse preprohypocretin is listed in SEQ ID NO: 2.

[0052] A hypocretin protein of this invention can be in a variety of forms, depending upon the use therefor, as described herein. For example, a hypocretin an be isolated from a natural tissue.

[0053] Alternatively, a hypocretin protein of this invention can be a recombinant protein, that is, produced by recombinant DNA methods as described herein. A recombinant hypocretin protein need not necessarily be substantially pure, or even isolated, to be useful in certain embodiments, although recombinant production methods are a preferred means to produce a source for further purification to yield an isolated or substantially pure receptor composition. A